Generation of Synthetic Benchmark Electrical Load Profiles Using Publicly Available Load and Weather Data

Gobind G. Pillai (gobind.pillai@northumbria.ac.uk), Ghanim A. Putrus (ghanim.putrus@northumbria.ac.uk), Nicola M. Pearsall (nicola.pearsall@northumbria.ac.uk)

Energy Systems Research Group, Engineering and Environment, Northumbria University, Ellison Building, Newcastle upon Tyne, NE1 8ST, UK.

Corresponding author:

Name: Professor Ghanim Putrus

Engineering and Environment, Ellison Building, Northumbria University

Newcastle upon Tyne, NE1 8ST, UK.

Email: ghanim.putrus@northumbria.ac.uk

Phone number: 00 44 (0) 191 2273107

Abstract

Electrical load profiles of a particular region are usually required in order to study the performance of renewable energy technologies and the impact of different operational strategies on the power grid. Load profiles are generally constructed based on measurements and load research surveys which are capital and labour-intensive. In the absence of true load profiles, synthetically generated load profiles can be a viable alternative to be used as benchmarks for research or renewable energy investment planning. In this paper, the feasibility of using publicly available load and weather data to generate synthetic load profiles is investigated. An Artificial Neural Network (ANN) based method is proposed to synthesize load profiles for a target region using its Typical Meteorological Year 2 (TMY2) weather data as the input. To achieve this, the proposed ANN models are first trained using TMY2 weather data and load profile data of neighbouring regions as the input and targeted output. The limited number of data points in the load profile dataset and the consequent averaging of TMY2 weather data to match its period resulted in limited data availability for training. This challenge was tackled by incorporating generalization using Bayesian regularization into training. The other major challenge was facilitating ANN extrapolation and this was accomplished by the incorporation of domain knowledge into the input weather data for training. The performance of the proposed technique has been evaluated by simulation studies and tested on three real datasets. Results indicate that the generated synthetic load profiles closely resemble the real ones and therefore can be used as benchmarks.

Keywords

Power networks, Synthetic generation, Load profile, Weather data, Artificial neural networks, Public data

1. Introduction

Power utilities across the world are proceeding with deregulation in order to facilitate system expansion by attracting investment from the private sector. As a consequence, in several countries (e.g. UK, India) the electricity sector has been unbundled into generation, transmission, distribution and supply companies. An electric load profile gives the dynamic variability of the energy demand with respect to time for the consumer category considered. The use of load profiles simplifies the arrangement between distribution companies and electricity suppliers, as the total energy record available in conventional energy meters can be distributed through different interval periods [1]. Having load profile information will lower the uncertainty in network management decision making for distribution companies [2]. It will also aid them in pricing competitive electricity options and devising risk hedging strategies for electricity derivatives.

There has been an increase in the deployment of renewable energy systems in recent years for many reasons such as the national and the global carbon emissions targets for combating climate change, growing electricity demand, the need for national energy security and providing energy access to populations without electricity. It would be prudent for the decision makers of developing nations to devise efficient future-minded policies and carefully plan their investment in renewables within their financial constraints. Information passed on to decision makers without the proper inclusion of hourly load dynamics can lead to cost-ineffective suboptimal energy systems which might also prevent countries reaching their set target of renewable energy use [3]. Load profiles are also essential for system operators to plan conventional (dispatchable) generation, to account for the energy import from renewable energy systems and for paying customers.

The standard method of constructing an hourly load profile is by recording the energy consumption, at feeder or substation level, at regular intervals (usually one hour) and dividing this by the number of customers on that feeder to produce an average demand (After Diversity Demand). Hourly load profile measurements at state-wise (sub regional) customer category levels are undertaken in developed nations, but are highly uncommon among developing nations like India owing to the capital requirement. Therefore, load profiling is accomplished by means of load research surveys whenever they are required. Surveying is a labour intensive process and it can take a long time to complete surveying for all states of a country. In many cases the results of these kinds of measurements or surveys when undertaken are archived in internal reports that are not easily accessible [4].

In the absence of true load profiles based on measurement or load research surveys, synthetically generated load profiles, which can be used as benchmarks, can be a viable alternative for research applications or renewable energy investment planning. In the past, predictive modelling of energy demand has been applied in a number of instances. Several studies have used ANN models for predicting energy consumption of individual and complex buildings [5, 6]. Aydinalp et al describe the modelling of the energy consumption in the Canadian residential sector using ANN models with weather and socio economic factors as inputs [5]. Magnano and Boland describe the synthetic generation of sequences of electricity demand for application in South Australia [7]. However, both these works have based their approaches on large datasets. The large data requirement makes them less easy to implement when the availability of data is limited, which is a typical characteristic of developing nations.

The load demand of a country or region depends on two major factors, the complexity of the economy and the weather of the area [8]. Studies have revealed that a large proportion of the variability in electricity demand is dependent on weather variables such as air temperature, humidity, wind speed, cloud cover and luminosity [9, 10]. The sensitivity of electricity demand in the commercial and residential sectors to meteorological variables is higher than in the industrial sector [11]. Both the weather and illumination components are very dependent on the hour of the day so they will have an impact on the daily load profile.

In the past, several artificial intelligence methods like ANNs, genetic algorithms, fuzzy logic, fuzzy expert systems, self-organising maps, wavelet transform, principal component analysis, grey system theory and support vector regression have been developed for forecasting electricity demand [8, 9, 12-14] Most of these methods are based on large datasets of historical time series of load data. In some cases, in addition to historical load data, both weather variables (temperature, relative humidity, wind velocity and cloudiness) and socio-economic factors have been used as inputs to the forecasting model [11]. Only recently, the synthetic generation of energy production or consumption using public data is starting to be recognized as a viable alternative to measurement and recording. Vladislavleva et al used genetic programming to demonstrate the feasibility of wind energy prediction by using publicly available weather and energy data for a wind farm in Australia [15].

In this paper, the feasibility of using publicly available datasets of load profile and weather data to synthetically generate load profiles using ANNs is investigated. A method is proposed to synthetically generate load profiles for the states of a developing nation like India, where measured or surveyed load profiles are not available, using only publicly available weather data of that state and the surveyed or measured load profiles of other neighbouring states. This is based on the assumption that the socio-economic tendencies of all states in that nation are similar. Bilgili et al employed a similar method for predicting a target station's wind speed using reference stations' data [16].

A general sketch of the proposed model is given in Figure 1. The developed ANN model is able to synthesize load profiles for a region using only TMY2 weather data. For achieving this, the model is first trained using load profile data of a region neighbouring the target region, with the targeted output and TMY2 weather data of this region as the input. The limited number of data points in the load profile dataset and the consequent averaging of TMY2 weather data to match its period resulted in limited data availability for training. This challenge was tackled by incorporating generalization using Bayesian regularization into training. The other major challenge was facilitating ANN extrapolation and was accomplished by the incorporation of domain knowledge into input weather data for training. It is to be noted that the validity of the benchmark load profiles generated synthetically are conditional and they are meant only to be used as a substitute for load profiling when the necessary load profile data is not available.

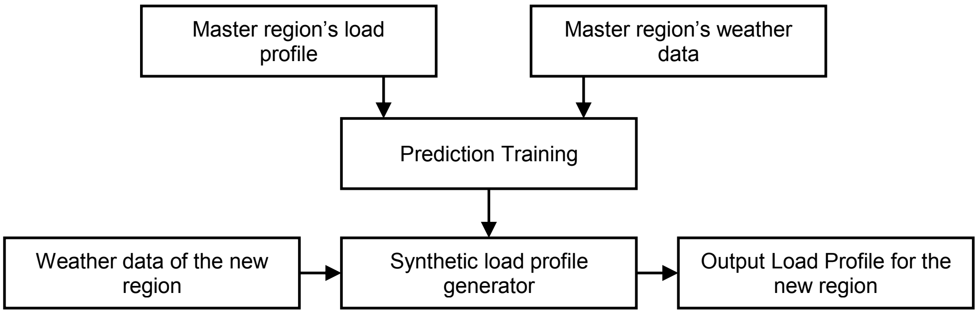


Figure 1: Proposed synthetic load profile generation method

This synthetic load profile generation also shares a common base with load forecasting in terms of the use of weather and economic factors; however, it does not make use of any mathematical load models or historical time series of load or weather datasets. There are three differences between the proposed method and forecasting: (1) typical meteorological year weather data is used as the input, (2) the load profile generated is for a region different from the training region (where measured data are available), (3) the load profile generated is in the same time frame as the training load profile rather than for the future.

The accuracy of the proposed technique has been evaluated by simulation case studies carried out on three real datasets. The first two case studies provide analysis and validation. In the third case study a modified version of the method is developed for application to a dataset based on research surveys. Results indicate that the synthetic load profiles generated closely resemble real ones and therefore can be used as benchmarks. A comparison of load profiles synthesised by ANN and regression demonstrate the superiority of ANN over regression for synthetic generation of load profiles.

The remainder of the paper is organized as follows: the datasets used in this study are described in section 2. In section 3, the formulation of the proposed synthetic load profile generation model is described. The feasibility of the method is investigated and validated by case studies in section 4. Finally, conclusions from this work are presented in section 5.

2. Data Description

2.1. Load profile data

Aggregate load profiles for Norway and Finland are taken from the Energy EFFiciency and LOad curve impacts of COMmercial development in competitive markets (EFFLCOM) project [17]. These load profiles represent the electrical energy consumption pattern at the national level and are based on system level measurements. The UK Energy Research Centre (UKERC) has developed UK load profile models for all seasons (separately), weekdays and weekends [18]. The load profile used in this study is Profile Class 1, which is based on hi-resolution metering at selected customer premises and applies to residential customers (unrestricted by usage timings). In Ireland, a customer smart meter electricity trial was undertaken in 2010 by the Commission for Energy Regulation (CER). Over 5,000 residential and commercial customers participated. The resulting data set which gives anonymised usage and behavioural patterns are available from the Irish Social Sciences Data Archive (ISSDA) [19]. A load profile was derived from the Irish data set for residential customers (which correspond to the UKERC's high summer residential load profile) by segregating the commercial customers' data. In India, the only load profile data available in the public domain are for the state of Gujarat, where a survey of about 400 residences was conducted to estimate the residential load profiles [20]. The survey covered all big and small cities/towns and the number of houses surveyed in any particular city was based on its share to the total population of the state.

Normalized values provide a useful mean to compare systems of different sizes. Therefore, normalized load profiles (as used in this paper) provide adequate information about the load values and profile information which allow comparison between countries of different sizes. Normalized load profiles are also valuable in understanding consumption behaviour of individual and group consumers. It can be the basis of interpreting occupancy and energy consumption during the hours of the day [21]. Normalizing the load profiles in per unit (p.u.) for all the case studies considered is based on dividing the actual value by the base power consumption. The base power consumption for the UK, Ireland and Gujarat is taken as 1 kW whilst load profiles available for Norway and Finland are normalized based on the average load of the day. For Norway and Finland, the average summer loads are 16.8 GW and 10.2 GW, respectively whilst the average winter loads are 11.8 GW and 8.2 GW, respectively.

2.2. Weather data

The US Department of Energy‘s (DOE) EnergyPlus software for Heating Ventilation Air Conditioning design uses meteorological data for energy flow calculations of residential and commercial buildings. Weather data for more than 2100 locations arranged by World Meteorological Organization region and country for a typical meteorological year are available from the DOE website [22]. The data are representative of the range of weather phenomena for the target location and result in annual averages that are consistent with the long term averages (from a data bank for a number of years) for that location. Norway, Finland, Ireland, United Kingdom and India were used as case study locations in this investigation. For India data, a set of 58 locations is available and is developed by ISHRAE (Indian Society of Heating, Refrigerating and Air Conditioning Engineers), an associate of ASHRAE (American Society of Heating, Refrigerating and Air Conditioning Engineers). For the European countries, data is available from IWEC (International Weather for Energy Calculation), which is an ASHRAE research project for 227 locations outside the US. The EnergyPlus TMY2 weather data files contain a number of data fields including hourly temperature, relative humidity, wind speed and cloud cover for all the months of a typical year. A complete list of the weather data fields and the weather data can be accessed from the DOE website [22].

All the load profiles and weather data described above have hourly time resolution. Typically there would be 2 load profiles or 1 load profile corresponding to the season (summer/winter) and there will be 12 sets of weather data corresponding to each month of a typical year. The weather data would then have to be averaged corresponding to the period covered by the load profile. This results in the amount of training data becoming limited. This is a major challenge as the synthetic generation model would have to learn from limited data, while maintaining its generalization and extrapolation capability.

3. Formulation of the Proposed Model

The most important component in the proposed model (see Figure 1) is the synthetic load profile generator which should be capable of learning from the training data and be able, based on the learning, to predict the output when new input data are presented. There are many such prediction mechanisms commonly used for engineering applications and, in this study, ANNs are used for predictive modelling.

ANNs are universal function approximators capable of mapping any nonlinear function. They are immune to error term assumptions and can tolerate noise, chaotic components, and extremities better than most other methods [23]. ANNs, like conventional regression analysis, attempt to minimize the sum of squared errors [24].

The recent abundance of literatures focusing on the application of ANNs for predicting energy consumption of individual and complex buildings and their applications in load forecasting indicates the superior capability of ANNs for predictive modelling [25].

3.1. Choosing the input variables of the ANN predictor: Sensitivity Analysis

The selection of input variables is a key issue in the design of ANN architectures and has been discussed in several studies [26-28]. It is determined ultimately by data availability [11]. The Energyplus dataset has many weather variables, any or all of which could be considered for inclusion in the ANN input vector. For N variables, there are 2^N-1 possible subsets to choose from and it is not usually practical to evaluate all subsets. It is desirable to rate input variables in terms of their importance in prediction within the context of a particular model. This is known as sensitivity analysis [29].

Several methods have been applied for sensitivity analysis of the input variables. Sousaa et al [1] used partial derivatives of each individual output with respect to each individual input, while Santos et al [28] employed the concepts of memory range (through block entropies estimation) and consumption tendency to define input vectors. Li et al observes that, in the field of building energy forecasting for buildings, Bayesian estimation, principle of maximum likelihood, statistical test for nonlinear correlation, etc. have been employed to detect relevant input variables [30]. Simplicity and easy data access aids the repeatability of the proposed synthetic load profile generation method. Most of the sensitivity analysis methods mentioned above require complicated analysis and are therefore not preferred.

In this work the R^2 statistic [31] is used for sensitivity analysis because it easily highlights an evident relation between the variables and the model. The R^2 statistic is determined for each weather data field (data set of a single variable) by generating a linear regression model with the load profile as the output and that particular data field as input. A low R^2 statistic does not indicate that there is no relationship between the weather data field and the load profile. It means that the ANN model has to be made more complex to capture the relationship. By eliminating such data fields in the input vector, the ANN model can be kept much more simple. Therefore, those weather data fields that satisfy R^2 =>0.1 for either or both summer and winter load profiles (the same predictor is used for both summer and winter load) are used in the ANN input vector.

3.2. Data Pre-processing

Data pre-processing is the process of analysing and transforming the input and output variables to minimize noise, highlight important relationships and detect trends [23]. In order to assist ANN learning, data pre-processing is usually performed by means of standardisation before ANN training. The standardisation of inputs and outputs that is required to run ANNs results in the minimum and maximum extremes in the training data set being scaled to the zero–one interval. Even if the use of a linear activation function in the output layer appears to offer a degree of extrapolation beyond fixed bounds, the confidence of predictions naturally deteriorates and confidence intervals become wider [15]. In practice the amount of extrapolation is limited by the saturation of the nodes in the hidden layer. Thus a problem arises in extrapolation; if the largest values in the testing data set exceed those in the training set, then the model may not provide good results.

Incorporating domain knowledge greatly enhances the confidence in the ANN model; prior transformation of the input data improves its generalisation capabilities without disrupting the training process or compromising its architecture [32]. Domain knowledge is added by scaling, using the formula

$$X_i = X_i - X_{global\ min} / (X_{global\ max} - X_{global\ min}) \quad (1)$$

where X_i is the i_{th} data point in the weather data field X of the region under consideration and $X_{global\ max}$ and $X_{global\ min}$ are respectively the largest and the smallest values in the global dataset of all regions considered for the weather data field X. The global set includes at least the weather data for the training region and the region for which the load profile is to be generated. Including weather data of more regions near to the training and required regions will facilitate better learning. The load data is unitised before using for training and therefore normally lies in the interval [0, 1].

3.3. ANN architecture

The ANN model complexity, the amount of data required and the difficulty in training the model increase as the number of elements in the ANN increases. An optimum design adequately models the problem with minimum network size [29]. For designing the ANN structure an approach similar to that described by Ochoa-Rivera et al was followed [33]. The target was to develop an optimum design with a simple architecture. A single hidden layer was considered between the input and output layers. Initially the ANN was trained with the same number of hidden nodes as output nodes. The ANN was then re-trained after reducing the number of hidden nodes to see if there was a significant improvement in ANN prediction performance. If there is a sufficient improvement, the new structure is chosen and the process of retraining and checking is repeated; otherwise the original architecture is maintained.

3.4. ANN training

Often, real world decisions are made with limited information. Despite the multitude of approaches researched in the field of artificial intelligence, most decisions rely on large amounts of data [34]. In the case of forecasting, reduction of training data limits the effectiveness of ANN-based forecasters as they become susceptible to over-fitting, where the model memorises the training data but fails to generalise well to new data [35]. When training data is limited, Chan et al describe five methods for improving generalization of ANN predictors [35]. They are the use of hyper parameters to each input variable, pruning and structural evolution, regularisation, cross validation and early stopping, and scaled conjugated gradient algorithm. Regularization was the method selected for this study owing to its ease of use and reproducibility. Generalisation using Bayesian regularization was incorporated into the Levenberg–Marquardt algorithm used to train the proposed ANN model [36, 37].

In the study of Sailor, state-wide monthly values of weather variables and consumer energy consumption data were used for energy demand forecasting [4]. The weather variables (temperature, humidity, and wind speed) were generated by population-weighting the regional climate data within each state. In a similar way, the data available for cities of a region was weighted by sample size from the load research survey to create average weather data for the state. In the case of measured load profiles, since they are not based on surveys, the weather data of the most populated city was taken as the average state data.

Once the ANN was trained, it was fed with weather data of the target region (pre-processed to incorporate domain knowledge), to synthetically generate the region's load profile. The entire synthetic generation procedure can be summarized as shown in Figure 2.

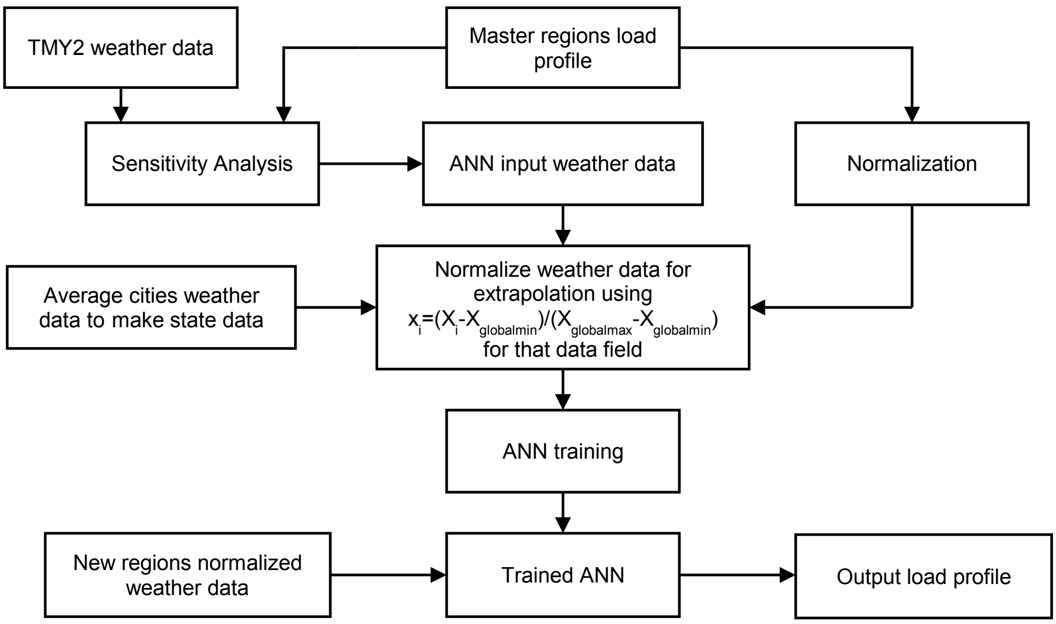


Figure 2: Proposed synthetic load profile generation procedure using ANN

4. Results and Discussion

In this section, the feasibility of the proposed method for synthetic load profile generation is investigated through three case studies. The results of the proposed ANN based technique are compared with the Multiple Linear Regression (MLR) based technique, as the latter is one of most popular prediction techniques. Statistical metrics are used to evaluate the learning ability of the ANN model and to measure the closeness of the synthetic benchmark load profile to the true load profile. The learning ability is evaluated by comparing the synthetic load profile generated when the training region's weather data is fed into the trained model to the training region's load profile. The closeness of synthetic load profile to the true load profile is evaluated by comparing the synthetic load profile generated by the ANN model against the actual surveyed or measured load profile of the particular region.

The statistical metrics examined used for examining the prediction accuracy are: the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE). Their definitions are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - A_i)^2}{n}} \qquad (2)$$

$$MAE = \frac{\sum_{i=1}^{n}|(P_i - A_i)|}{n} \qquad (3)$$

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{(P_i - A_i)}{A_i}\right|}{n} * 100 \qquad (4)$$

Where P_i and A_i are the synthetic load profile data and actual load profile data at the i^{th} point respectively, and n is the total number of data points (i.e. 24 per load profile for hourly resolution).

4.1. Model Validation

4.1.1. Case study 1: Aggregate load profiles

In this case study, Finland is considered as the region for which aggregate synthetic load profiles for summer and winter are to be generated and Norway is considered as the training region. Both countries are in the NORDIC region and have similar social and economic patterns and follow similar climatic regimes. Therefore, they are macro-social units that can be compared in socio-economic and meteorological terms. Energyplus weather data files are available for the major cities of Helsinki and Oslo.

4.1.1.1. Input preparation

The global weather data set for the study included weather variables common to Norway and Finland. A sensitivity analysis was performed on Norway's global weather variables to its summer and winter aggregate load profile (the total load at the system level normalized by the average hourly load) [17]. The results of the sensitivity analysis are shown in Table 1. The weather variables that meet the criterion for sensitivity (as mentioned in section 3) were selected and normalized using eq.1 to form the ANN input vector. The summer and winter weather datasets were the input vectors for the ANN model training, while the corresponding summer and winter load profiles were the targeted training outputs.

Table 1: Norway weather variable sensitivity

R2 Coefficient for:	**Norway**		
	Winter	**Summer**	**Combined**
Temperature	0.51	0.62	0.08
Direct Normal	0.27	0.64	0.35
Humidity	0.50	0.64	0.26
Wind speed	0.59	0.55	0.43
Opaque Sky cover	0.05	0.21	0.01

4.1.1.2. ANN design and training

The ANN architecture used in this case study is shown in Figure 3. Initially, training was performed using the Levenberg–Marquardt training algorithm. Later the Levenberg–Marquardt training algorithm was modified to include Bayesian Regularization [37] targeting maximum learning from the same datasets.

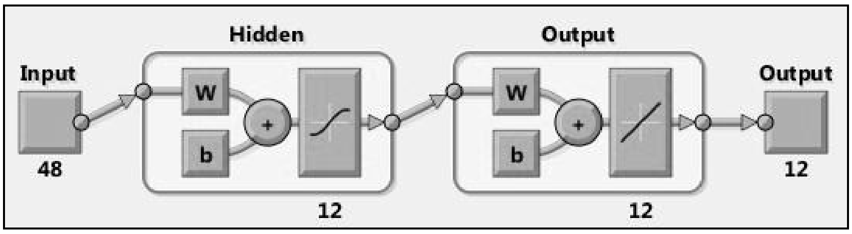
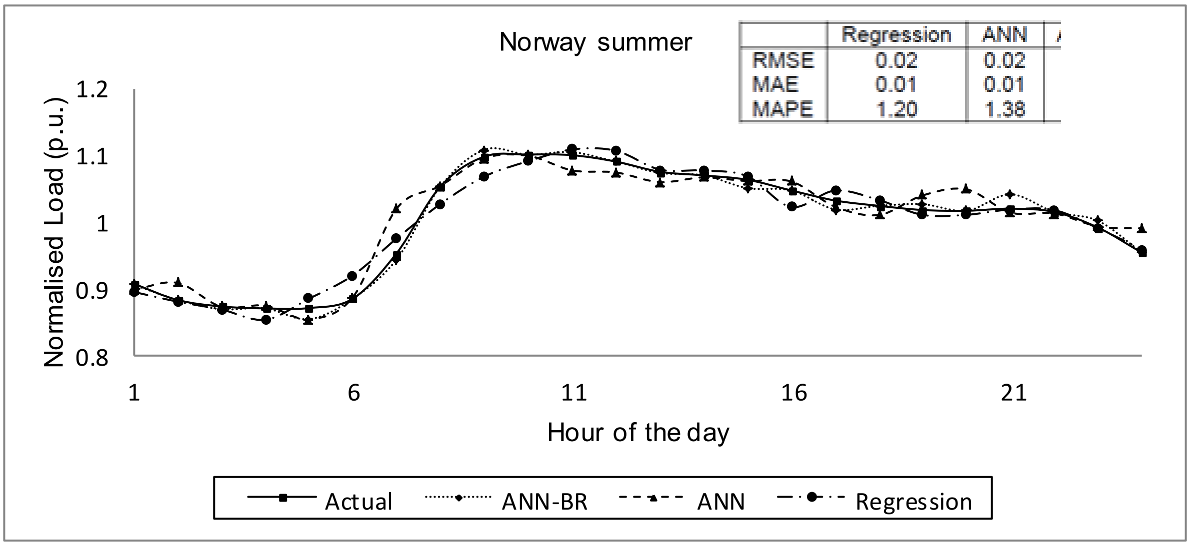


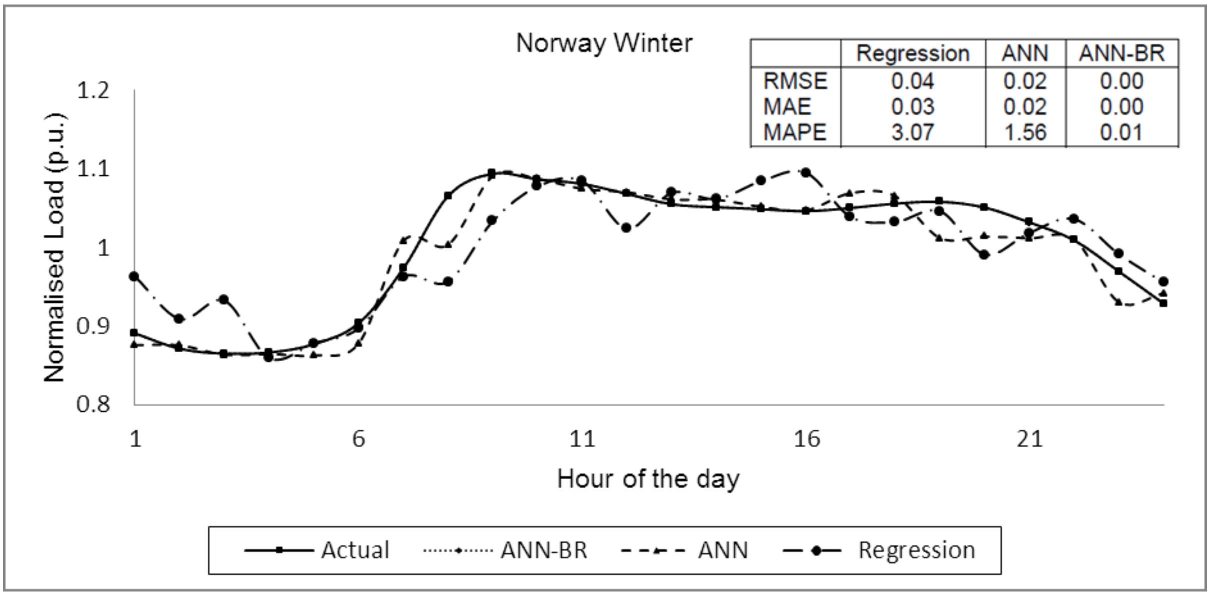
Figure 3: ANN architecture for NORDIC case study

4.1.1.3. Synthetic load profile generation

Finland's normalized weather data was fed to the trained ANN to generate its synthetic aggregate summer and winter load profiles. In Figure 4, Norway's actual load profile is compared with the synthetic load profile generated from the proposed ANN model (using the training weather data). It can be seen from the prediction metrics which are embedded in the figure that the ANN technique outperforms the MLR (regression) method. The ANN learning enhancement when Bayesian regularization is included in the training algorithm can also be seen. Figure 5 shows the synthetic load profiles generated for Finland together with the real measured load profiles. It can be seen from the prediction metrics that the synthetically generated load profiles closely resemble the real load profiles.

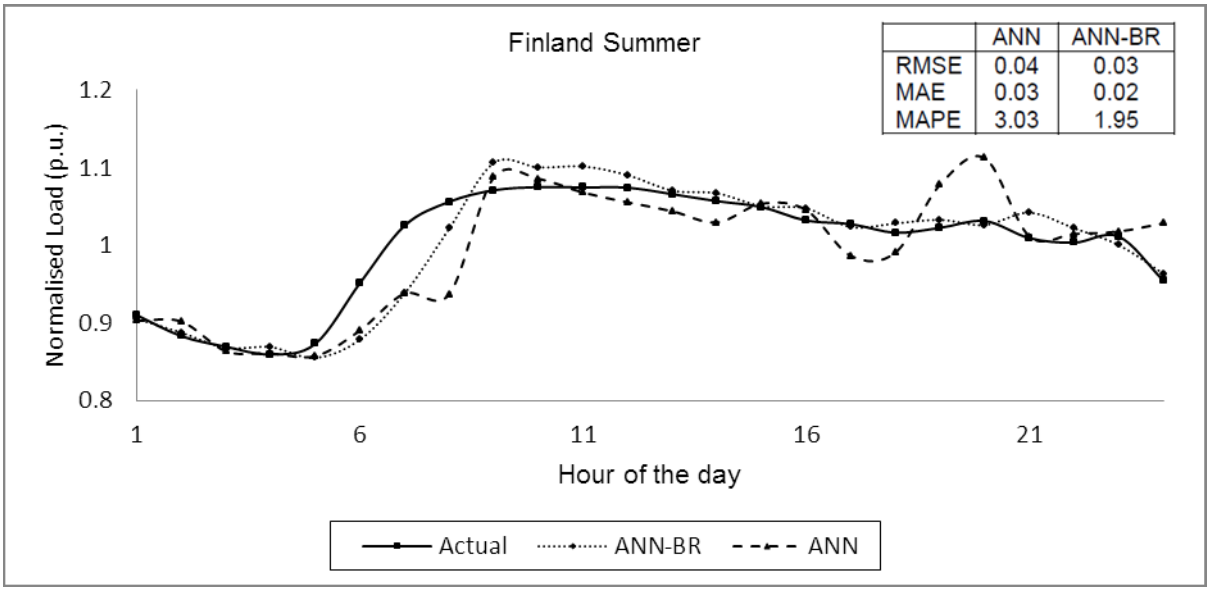


(a) Summer load profiles

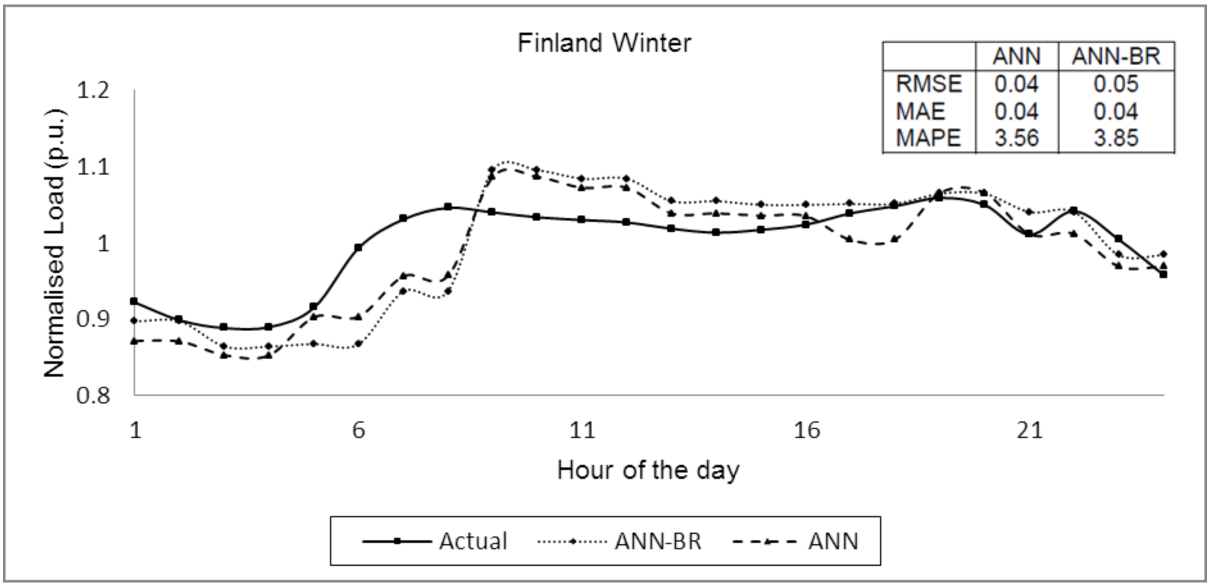


(b) Winter load profiles

Figure 4: Norway aggregate load profiles and prediction metrics



(a) Summer load profiles



(b) Winter load profiles

Figure 5: Finland aggregate load profiles and prediction metrics

4.1.2. Case study 2: Residential load profiles

In this case study, the UK is considered as the region for which residential synthetic load profiles for 'high summer' are to be generated and Ireland is considered as the training region. According to UKERC, the 'high summer' is defined as the period of six weeks and two days from the sixth Saturday before August Bank Holiday up to and including the Sunday following August Bank Holiday [18]. As in the previous case, both countries have similar social and economic patterns and follow similar climatic regimes which make them comparable in socio-economic and meteorological terms. Energyplus weather data files are available for the major cities of London and Dublin.

4.1.2.1. Input preparation

A high summer residential load profile for Ireland was derived from the Irish dataset [19]. The global weather data set for the study included weather variables common to Ireland and the UK. A sensitivity analysis was performed on Ireland's global weather variables in relation to its high summer load profile and the results are shown in Table 2. The ANN input vector is formed as described previously. For ANN training, the high summer weather data set was the input and the corresponding load profile was the targeted training output.

Table 2: Ireland weather variable sensitivity

R2 Coefficient for:	**Ireland HSR**
Temperature	0.68
Direct Normal Irradiance	0.32
Humidity	0.69
Opaque Sky cover	0.00

4.1.2.2. ANN design and training

The ANN training followed the same sequence as in the first case study. The ANN architecture used is as shown in Figure 6.

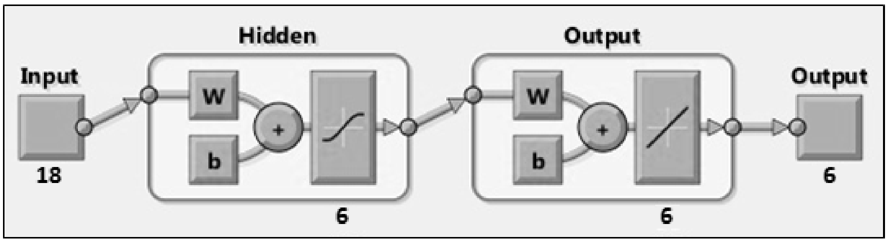


Figure 6: ANN architecture for Ireland-UK case study

4.1.2.3. Synthetic load profile generation

Once the ANN model was trained, the UK's normalized high summer weather data was fed to the ANN to generate its synthetic residential high summer load profile. As in the previous case study, Figure 7 shows Ireland's actual load profile together with the synthetic load profiles generated from the proposed ANN model (using training weather data).The results show the superiority of the ANN technique over the MLR method and the ANN learning enhancement when Bayesian regularization was included in the ANN training algorithm.

Figure 8 shows the synthetic load profiles generated for UK together with the actual load profiles of UK from UKERC. It can be seen from the prediction metrics that the synthetically generated load profile closely resembles the actual load profile.

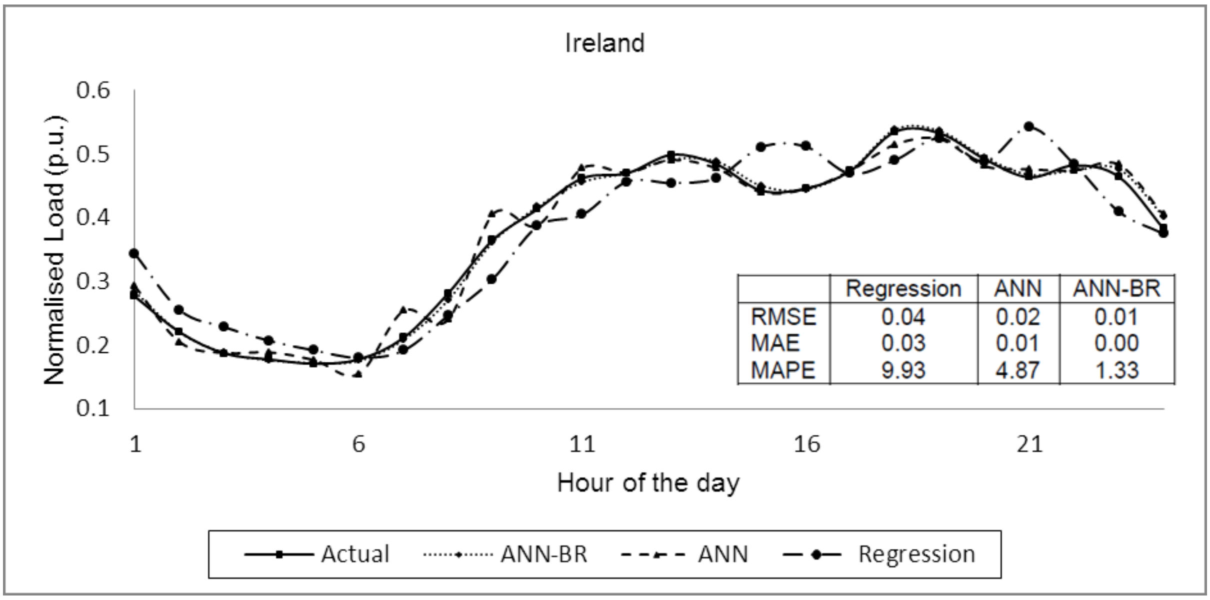


Figure 7: Ireland residential high summer load profile training results and prediction metrics

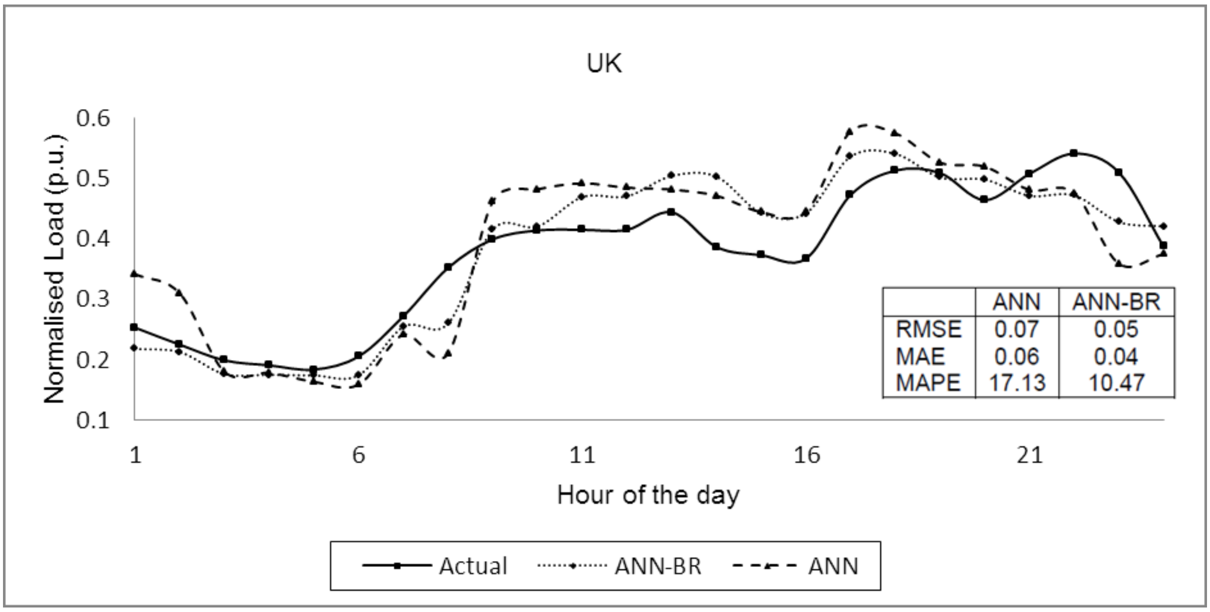


Figure 8: UK residential high summer load profile prediction results and metrics

The two case studies described earlier demonstrate that it is feasible to obtain a satisfactory prediction of synthetic benchmark load profiles, even when using limited public datasets using the proposed method. The results obtained also show that the ANN technique, because of its advance learning capabilities, outperforms MLR (regression) in the role of predictor for the proposed model.

4.2. Application of the proposed method

4.2.1. Case study 3: Residential load profiles for Indian states

EnergyPlus weather data files are available for most major cities in developing countries. However, residential load profiles are rarely available in the public domain, e.g. literature shows data available for only one Indian state. According to Rallapalli and Ghosh, scientific estimations of electricity demand in India are not common [38]. As there are not enough real data from Indian states to test its viability, the previous two case studies were considered as necessary and sufficient validation to the proposed approach. If the states of India are considered as macro-social units (a geographic region representative of a large population), using the proposed method an ANN trained on data from one state could be used for predicting the benchmark load profiles of a neighbouring state.

The success of the proposed method depends on the learning ability of the ANN model. An ANN model was trained with the normalized average weather data and the summer and winter residential load profiles of the state of Gujarat. As in the previous cases, the actual load profile was compared with the synthetic load profile (fed with real weather data). The MAPE for Norway was 0.48% and 0.01%, while for Ireland it was 1.33%. In this case for Gujarat, the prediction metrics indicated that the fit to training data was not satisfactory (MAPE was 9.06%). Therefore, the methodology was modified based on a Gujarat Load Research survey, which gave the load usage categorized in terms of lighting, space-cooling etc. as follows:

- Gujarat's summer and winter domestic load profiles were segregated into *Lighting*, *Space-cooling* and '*Others*'. The '*Others*' load profile is mainly dependant on socio-economic factors. It was considered to be the same for the training region and the target region.
- The synthetic load profile generation methodology described previously was applied separately on both *Lighting* and *Space-cooling* load profiles resulting in the generation of synthetic *Lighting* and synthetic *Space-cooling* load profiles.
- The target regions load profile was then the sum of the synthetic *Lighting,* synthetic *Space-cooling* and the original '*Others*' load profiles.

Input preparation and ANN design and training were conducted as described in the previous case studies. Table 3 shows the sensitivities of the *Lighting* and *Space-cooling* load profiles to weather variables. Inputs to the corresponding ANN models were chosen based on the sensitivities.

Table 3: Gujarat weather variable sensitivity

R2 Coefficient for:	**Lighting Load**			**Space Cooling**		
	Winter	**Summer**	**Combined**	**Winter**	**Summer**	**Combined**
Temperature	0.13	0.07	0.07	0.29	0.37	0.21
Direct Normal Irradiance	0.24	0.24	0.26	0.32	0.39	0.27
Humidity	0.10	0.04	0.06	0.23	0.33	0.18
Opaque Sky cover	0.02	0.00	0.02	0.04	0.00	0.05

As can be noticed from Tables 1-3, there is quite a variation in the results of R^2 coefficient for different variables and different regions. This can be best explained by taking a closer look at Table 3. It can be seen that the S*pace-cooling* load has a higher R^2 coefficient for Summer time temperature and Direct Normal Irradiance (DNI). During summer, there is a higher solar input in Gujarat and consequently a higher air-conditioning demand. It is evident from the literature [8-11] that the relation between the weather variables and the load profile is location and season specific. The R^2 coefficients have different values, which depend on the relation of that weather variable to the load profile. As explained in Section 3.1, an R^2 coefficient < 0.1 does not indicate that there is no relationship between the weather data field and the load profile. The ANN model can be simplified by eliminating such data fields, else the ANN model becomes too complex to enable capturing the relationship.

Table 4 provides a summary of weather variables selected and data sample size. At an hourly resolution, the number of data points per data field (M) is 24.

Table 4: Summary of weather variables selected and data sample size

Case study	Training Region	Load Profile data fields	Selected weather variable data fields	Sample size per data field
1	Norway	Summer and Winter aggregate load profiles	Temperature, DNI, Humidity, Windspeed	2M
2	Ireland	High summer residential load profiles	Temperature, DNI, Humidity	M
3.a	Gujarat	Summer and Winter residential load profiles	Temperature, DNI, Humidity	2M
3.b	Gujarat	Summer and Winter residential lighting/space-cooling load profiles	Temperature, DNI, Humidity	2M

Figure 9 and 10 shows the residential load profiles generated synthetically using the original methodology and the modified methodology respectively. The prediction metrics depict that a satisfactory fit was obtained to the training load profile using this modified method. A better fit to input training data was obtained when category load profiles were used because the ANN develops better relations in this case. Load profiles for other neighbouring Indian states could be generated by inputting the weather data to the *Lighting* and *Space-cooling* ANN synthetic load profile predictors and summing the outputs with the original '*Others'* load profile.

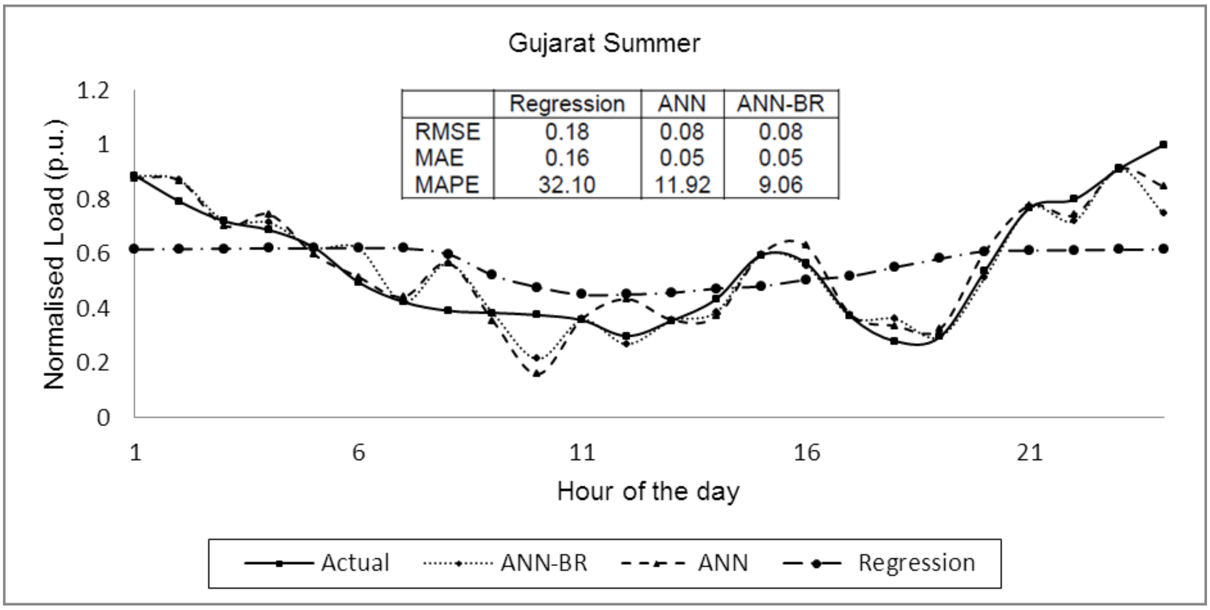


Figure 9: Gujarat summer load profile results and prediction metrics using the original approach.

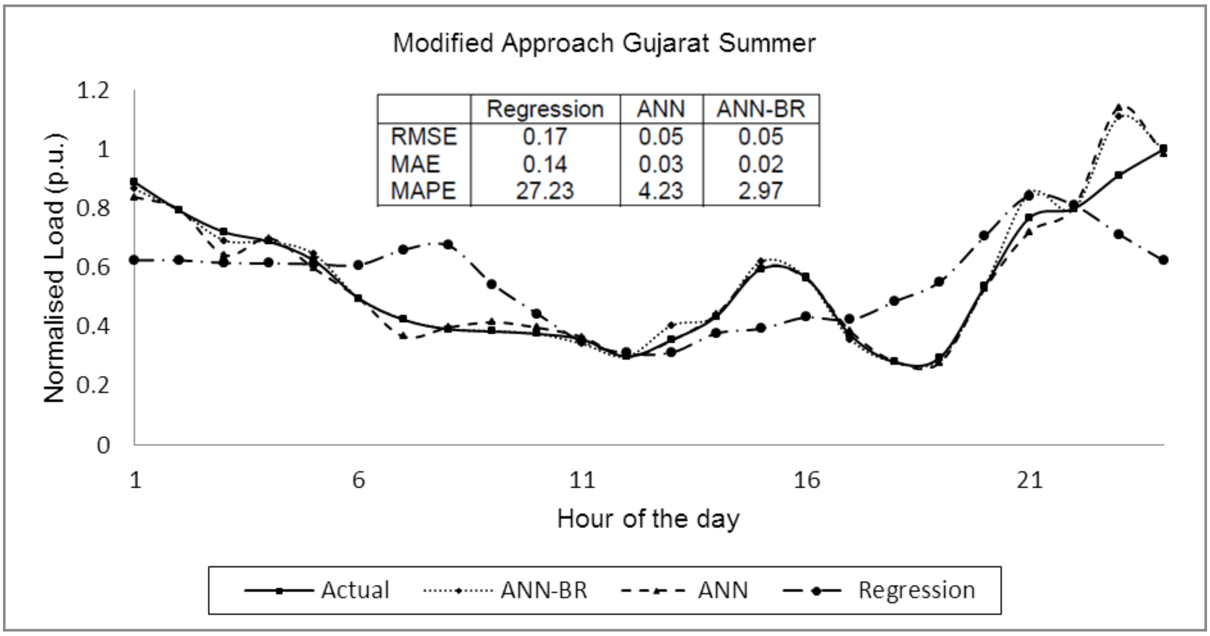


Figure 10: Gujarat summer load profile results and prediction metrics using the modified approach.

From the above case studies the comparison of the results from the ANN model trained with Levenberg–Marquardt, with and without the inclusion of Bayesian Regularization, demonstrates the simultaneous error reduction and achievement of greater generalization when Bayesian Regularization is included in the training algorithm. Note that regression results are presented only for training regions. For target regions they tended to generate negative load profiles and are therefore not shown.

5. Conclusions

A simple ANN based technique for generating synthetic load profiles using limited publicly available load profiles and weather data has been developed. The synthetically generated load profiles have the same diurnal characteristics as actual load profiles and also show the same correlation to weather and other parameters that affect energy consumption. Both a direct validation through comparison of synthetically generated load profiles to actual load profiles, as well as an indirect validation through synthetic load predictor trainability, show that the proposed method of synthetic generation is simple, reproducible and robust in learning from limited real data. The limitations of practically available load profile data and how the proposed method can be adapted to address these limitations were identified in case study 3. The most important feature of the proposed method is functionality, as it allows realistic benchmark load profiles to be generated for locations which would otherwise rely on other inaccurate methods, such as a constant load assumption. The only limitation is the availability of weather and load data for certain locations, which is required for model training.

This research work is continuing to investigate the potential of extending this method to create daily regional synthetic load profiles based on daily regional weather data. The incorporation of socio-economic factors to the synthetic load profile generation technique is also an area that merit further investigation. Probability distributions of socio-economic factors which affect the load, such as the regional GDP figures or schedule of working hours, could be used to scale the load model.

Acknowledgment

This work has been supported by a joint UK-India initiative in solar energy through a joint project ‘Stability and Performance of Photovoltaics (STAPP)’ funded by Research Councils UK (RCUK) Energy Programme in UK (contract no: EP/H040331/1) and by the Department of Science and Technology (DST) in India. We would like to thank the Irish Social Science Data Archive (ISSDA), for providing us the CER Electricity customer behaviour trial data set.

References

[1] Sousaa JC, Nevesa LP, Jorgeb HM. Assessing the relevance of load profiling information in electrical load forecasting based on neural network models. International Journal of Electrical Power & Energy Systems. 2012;40:85-93.

[2] ELEXON. Load profiles and their use in electricity settlement; 2013. < www.elexon.co.uk/ELEXON%20Documents/load_profiles.pdf >.

[3] Pinaa A, Silvac C, Ferrãoa P. Modeling hourly electricity dynamics for policy making in long-term scenarios. Energy Policy. 2011;39:4692-702.

[4] Sailor DJ. Relating residential and commercial sector electricity loads to climate-evaluating state level sensitivities and vulnerabilities. Energy. 2001;26:645-57.

[5] Aydinalp M, Ugursal VI, Fung AS. Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. Applied Energy. 2002;71:87-110.

[6] Kreider JF, Claridge DE, Curtiss P, Dodier R, Haberl JS, Krarti M. Building energy use prediction and system identification using recurrent neural networks. Journal of Solar Energy Engineering. 1995;117:161-6.

[7] Magnano L, Boland JW. Generation of synthetic sequences of electricity demand: Application in South Australia. Energy. 2007;32:2230-43.

[8] Bunnoon P, Chalermyanont K, Limsakul C. Mid-Term Load Forecasting: Level Suitably of Wavelet and Neural Network based on Factor Selection. Energy Procedia. 2012;14:438-44.

[9] Wang C-H, Grozev G, Seo S. Decomposition and statistical analysis for regional electricity demand forecasting. Energy. 2012;41:313-25.

[10] Thatcher MJ. Modelling changes to electricity demand load duration curves as a consequence of predicted climate change for Australia. Energy. 2007;32:1647-59.

[11] Apadula F, Bassini A, Elli A, Scapin S. Relationships between meteorological variables and monthly electricity demand. Applied Energy. 2012;98:346-56.

[12] Amina M, Kodogiannis VS, Petrounias I, Tomtsis D. A hybrid intelligent approach for the prediction of electricity consumption. International Journal of Electrical Power & Energy Systems. 2012;43:99-108.

[13] Santana ÁL, Conde GB, Rego LP, Rocha CA, Cardoso DL, Costa JCW, et al. PREDICT – Decision support system for load forecasting and inference: A new undertaking for Brazilian power suppliers. International Journal of Electrical Power & Energy Systems. 2012;38:33-45.

[14] Hong W-C, Dong Y, Zhang WY, Chen L-Y, K. Panigrahi B. Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm. International Journal of Electrical Power & Energy Systems. 2013;44:604-14.

[15] Vladislavleva K, Friedrich T, Neumann F, Wagner M. Predicting the Energy Output of Wind Farms Based on Weather Data: Important Variables and their Correlation. Renewable Energy. 2013;50:236-43.

[16] Bilgili M, Sahin B, Yasar A. Application of artificial neural networks for the wind speed prediction of target station using reference stations data. Renewable Energy. 2007;32:2350-60.

[17] EFFLOCOM. Load Profiles. 2011. < http://www.sintef.no/Projectweb/Efflocom/DisseminationDocumentation/ >.

[18] UKERC. Electricity user load profiles by profile class. 2011. < http://data.ukedc.rl.ac.uk/cgi-bin/dataset_catalogue/view.cgi.py?id=6 >.

[19] CER. Electricity Customer Behaviour Trial. 2011.

[20] Garg A, Maheshwari J, Upadhyay J. Load Research for Residential and Commercial Establishments in Gujarat. Energy Conservation and Commercialization. 2010. < http://eco3.org/load-research-report-for-gujarat-1024/ >.

[21] Kwac J, Flora J, Rajagopal R. Household energy consumption segmentation using hourly data. Smart Grid, IEEE Transactions on. 2014;5.1: 420-430.

[22] US DoE. Weather Data, Energy Efficiency & Renewable Energy, Building Technologies Program; 2012. < http://apps1.eere.energy.gov/buildings/energyplus/weatherdata_about.cfm >.

[23] Kaastra I, Boyd M. Designing a neural network for forecasting financial and economic time series. Neurocomputing. 1996;10:215-36.

[24] Akole M, Bongulwar M, Tyagi B. Predictive model of load and price for restructured power system using neural network. In: Proceedings of 2011 International Conference on Energy, Automation and Signal (ICEAS); 2011. p. 1-6.

[25] Suganthi L, Samuel AA. Energy models for demand forecasting—A review. Renewable and Sustainable Energy Reviews. 2012;16:1223-40.

[26] Fidalgo JN, Lopes JAP. Load forecasting performance enhancement when facing anomalous events. IEEE Transactions on Power Systems, . 2005;20:408-15.

[27] Ramezani M, Falaghi H, Haghifam M, Shahryari GA. Short-term electric load forecasting using neural networks. In: Proceedings of the International conference on computer as a tool EUROCON; 2005. p. 1525-8.

[28] Santos PG, Martins A, Pires AJ. Designing the input vector to ANN-based models for short-term load forecast in electricity distribution systems. International Journal of Electrical and Power Energy Systems. 2007;29:338–47.

[29] Hunter A, Kennedy L, Henry J, Ferguson I. Application of neural networks and sensitivity analysis to improved prediction of trauma survival. Computer Methods and Programs in Biomedicine. 2000;62:11-9.

[30] Li K, Su H, Chu J. Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study. Energy and Buildings. 2011;43:2893-9.

[31] Colorado Uo. Statistics Overview: Regression Analysis. 2012. < www.colorado.edu/eeb/courses/2040/Lab_Data/documents/StatisticsOverviewRegression.pdf >.

[32] Hettiarachchi P, Hall MJ, Minns AW. The extrapolation of artificial neural networks for the modelling of rainfall–runoff relationships. Journal of Hydroinformatics. 2005;7:292-6.

[33] Ochoa-Rivera JC, García-Bartual R, Andreu J. Multivariate synthetic streamflow generation using a hybrid model based on artificial neural networks. Hydrology and Earth System Sciences. 2002;6:641-54.

[34] Chang FM. An intelligent method for knowledge derived from limited data. In: IEEE International Conference on Systems, Man and Cybernetics; 2005. p. 566-71.

[35] Chan ZSH, Ngan HW, Rad AB, David AK, Kasabov N. Short-term ANN load forecasting from limited data using generalization learning strategies. Neurocomputing. 2006;70:409-19.

[36] Mathworks. Matlab R2012b documentation: Levenberg-Marquardt backpropagation; 2012. < http://www.mathworks.co.uk/help/nnet/ref/trainlm.html >.

[37] Souza CR. Neural Network Learning by the Levenberg-Marquardt Algorithm with Bayesian Regularization; 2012. < http://crsouza.blogspot.co.uk/2009/11/neural-network-learning-by-levenberg_18.html >.

[38] Rallapalli SR, Ghosh S. Forecasting monthly peak demand of electricity in India—A critique. Energy Policy. 2012;45:516-20.